TO WHAT EXTENT DO BULL AND BEAR MARKET CONDITIONS IMPACT THE EFFECTIVENESS OF CERTAIN FEATURES IN FORECASTING BITCOIN PRICE?

MARIJN VAN WATERSCHOOT

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
AT THE SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
OF TILBURG UNIVERSITY

STUDENT NUMBER

2028787

COMMITTEE

dr. Gonzalo Nápoles
dr. Maryam Alimardani

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

January 15th, 2024

WORD COUNT

8789

ACKNOWLEDGMENTS

I am immensely grateful to Dr. Gonzalo Nápoles for his invaluable guidance and support throughout my research. His mentorship has been crucial to my work. I also wish to thank my family and friends for their continuous encouragement and belief in me. Their support has been a pillar of strength in my endeavors.

TO WHAT EXTENT DO BULL AND BEAR MARKET CONDITIONS IMPACT THE EFFECTIVENESS OF CERTAIN FEATURES IN FORECASTING BITCOIN PRICE?

MARIJN VAN WATERSCHOOT

Abstract

Bitcoin, a pioneering cryptocurrency, has captivated global interest with its potential to revolutionize financial transactions. Its market, marked by high volatility, presents unique challenges and opportunities for investors and analysts. This thesis explores the impact of bull and bear market conditions on the effectiveness of in forecasting Bitcoin prices. The research investigates how these market conditions affect the use of historical and sentiment data in predicting Bitcoin prices. This study uniquely focuses on feature effectiveness variability across market phases, an aspect not thoroughly examined in prior research. The dataset comprises Bitcoin price data from Yahoo Finance and related sentiment data from Augmento.ai. The methodology employs Long Short-Term Memory and Gated Recurrent Unit models, alongside the ARIMAX model, for the prediction of Bitcoin prices. Subsequently, SHapley Additive exPlanations are incorporated for the feature importance analysis. The findings of this study indicate that financial features are more influential than sentiment features in both bull and bear markets, particularly in bear markets. The research contributes to a nuanced understanding of market dynamics and feature influence in cryptocurrency forecasting, especially across varying market conditions.

1 DATA SOURCE, ETHICS, CODE, AND TECHNOLOGY STATEMENT

The data utilized in this thesis comprises two distinct datasets. The first dataset, consisting of Bitcoin USD (BTC-USD) data, has been acquired from Yahoo Finance. This dataset was accessed via Yahoo Finance's publicly available interface and is used in compliance with their data use policies.

The second dataset, pertaining to sentiment data related to Bitcoin, was obtained from Augmento.ai. They provided this dataset, which is freely available and comes in a CSV format. This dataset is used under the terms and conditions provided by Augmento.ai and is in line with their data usage guidelines. This thesis did not involve the collection of data from human participants or animals. Both Yahoo Finance and Augmento.ai retain ownership of their respective datasets during and after the completion of this thesis. Ethical approval was not required since the data is publicly available and does not contain sensitive personal information. All figures created for this thesis are the original work of the author. The code utilized in the thesis is primarily original. In the process of writing and coding for this research, ChatGPT, a generative language model, was used for assistance in debugging, code optimization, language paraphrasing, spell checking, and grammar enhancement. This usage complies with OpenAI guidelines for ChatGPT. No other typesetting tools or external writing services were engaged in the preparation of this thesis.

2 INTRODUCTION

2.1 *Motivation*

After the introduction of Bitcoin in 2008, it gained immense popularity, evolving from an obscure digital token to a financial powerhouse. Bitcoin promised a future where transactions could be made without the intervention of traditional banking systems. Cryptocurrencies, in general, have become so popular that even the United States is considering introducing its own digital currency. Joe Biden has signed an executive order addressing government oversight of cryptocurrency, encouraging the Federal Reserve to consider creating its own digital currency (Guardian, 2022). Recent years have seen an increase in investment activities in cryptocurrencies. Investing in cryptocurrencies is becoming increasingly common, especially among young people (NOS, 2022). In November 2021, the cryptocurrency reached a market capitalization of roughly $3 trillion. Shortly after the peak, the market crashed to $900 billion (Ari & MacKenzie, 2022). These significant price swings show that while gains are attainable, substantial losses are also possible. The extreme volatility makes it very important to understand the market dynamics. This is particularly important as young, vulnerable individuals often make these investments. Accurate forecasts can guide cryptocurrency investors in making informed investment decisions, potentially resulting in enhanced returns. Additionally, such predictions could also be helpful for researchers studying the market behavior (Pintelas et al., 2020).

Current research provides valuable insights into predicting Bitcoin prices using historical prices and sentiment features (Pathak & Kakkar, 2020). however, there remains a gap in understanding how these features behave during different market phases. Understanding which indicators hold the most influence in diverse market conditions holds a strong societal importance as it can impact the financial literacy and education of the public. Educating the public about market dynamics, including the differences between bull and bear markets, can empower individuals in making better financial decisions and understanding the risks associated with cryptocurrency investments.

2.2 *Project definition*

The previous motivation leads to the following research question:

> **Main Research Question:** To what extent do bull and bear market conditions impact the effectiveness of historical data and sentiment data in forecasting Bitcoin price?

The main research question aims to investigate the influence of market conditions, specifically bull and bear phases, on the efficacy of using historical and sentiment data to predict the price of Bitcoin. The goal is to assess whether these forecasting techniques are effective in both bull and bear markets, or if specific market conditions enhance or reduce their forecasting strength. To further unpack the primary research question, multiple sub-questions have been formulated and outlined below:

> **SQ1**. How effectively can the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models predict Bitcoin's value based solely on historical data compared to the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model?

Before introducing other features, it is important to determine the base predictive power of the LSTM and GRU models using just the historical price data of Bitcoin. These models are compared to the baseline model (ARIMAX). This comparison validates that the LSTM and GRU benefits are not merely random outcomes, but signify true advancements over conventional, more straightforward prediction techniques. This sets a reference point that can later be compared to a model with additional features. The models will be compared not only on Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) metrics but also through visualization of their predictions, focusing on identifying any potential x-axis shifting. The dataset employed for this analysis covers the period from November 1, 2016, to June 18, 2023.

SQ2 To what extent do supplementary sentiment features influence the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) of the LSTM and GRU models compared to the ARIMAX model?

This question assesses the incremental value of other features beyond the historical data. Examining how the LSTM, GRU, and ARIMAX models change with the inclusion of additional features will gain insights into their usefulness in forecasting the Bitcoin price.

SQ3. Which features emerge as the most influential when conducting a SHAP feature importance analysis on the best performing model during bull and bear market phases?

In addressing the third research question, the study will segment the data according to bull and bear markets identified by the Simple Moving Average. By analyzing which features are most influential during specific market conditions using SHapley Additive exPlanations (SHAP), it can be determined how these conditions impact the effectiveness of each metric.

3 RELATED WORK

This section reviews literature on Bitcoin price prediction, methodically structured. It starts with key determinants identified in prior studies, analyzing historical price data, sentiment indicators from social media and news, and technical indicators. Next, it evaluates the effectiveness of various predictive models used in these studies for price forecasting. The discussion then moves to Feature Importance Analysis for deeper insights. Finally, it synthesizes these findings to pinpoint the research gap, highlighting the current knowledge in cryptocurrency price prediction and areas needing further research.

3.1 *Forecasting features*

A significant amount of research has been conducted on predicting cryptocurrency prices, reflecting the growing interest and importance of digital assets in the financial landscape. Multiple studies have indicated that historical price data can effectively forecast future prices of cryptocurrencies. In a 2021 study, Hari Krishnan Andi shows that with a large dataset, the price of Bitcoin can be efficiently predicted with accurate results (Andi, 2021). Another study demonstrated that historical data could forecast the prices of Zcash and Litecoin 1 day, 3 days, and 7 days ahead with notable accuracy (Patel et al., 2020). Using data points like daily average price, opening price, highest and lowest values, and trading volume, the study highlighted the success of a deep learning approach in capturing the complex movements of cryptocurrency markets.

In a different research investigation, Wu et al. (2018) used a small dataset with only 209 instances of daily price data and transaction volumes to analyze trends and patterns. Through this detailed examination, the study provides insights into the fluctuating nature of Bitcoin prices and their relationship with transaction volumes, offering a nuanced perspective on the economic factors influencing the cryptocurrency market.

Other studies show that sentiment expressed in news articles and social media platforms plays a pivotal role in shaping the dynamics of cryptocurrency prices. Research shows that public perception and behavioral attitudes often affect cryptocurrency pricing (Wołk, 2020). In this study, Wołk found that the short-term price swings are substantially influenced by the sentiment on social media. It concludes that social media sentiments are crucial in predicting short-term cryptocurrency market movements, illustrating the intricate relationship between public opinion and cryptocurrency dynamics.

Research by Valencia et al. (2019) demonstrates that machine learning, coupled with sentiment analysis – particularly when applied to Twitter data – can effectively forecast cryptocurrency markets using neural networks. The study aims to predict the price movement of cryptocurrencies, focusing on Bitcoin, Ethereum, Ripple, and Litecoin. The combination of market data and sentiment analysis derived from social media offers a more nuanced and effective approach for predicting price movements in cryptocurrencies. However, the study also showed that while Twitter data alone was not sufficient for prediction, it contributed to improved precision in certain cases. This is also mentioned in another study by Inamdar et al. (2019). In this study Inamdar et al. reveals that sentiment scores from social media do not significantly impact Bitcoin prices unless these scores are biased towards a particular sentiment. The author emphasizes that historical data and Bitcoin's trading volume are the main factors in predicting its price. A study by Abraham et al. (2018) shows that even the volume of Twitter messages and Google Trend data can be used to accurately predict the direction of price changes.

There are also multiple studies that used a combination of both historical and sentiment data. Combining features from social media with historical pricing data yields effective results, as demonstrated in the (Pathak & Kakkar, 2020) study on cryptocurrency. This study employed tools like Tweepy and GetOldTweets for gathering Twitter data and utilized TextBlob for conducting sentiment analysis. The findings of the paper suggest that the integration of insights derived from social media sentiment with historical market data proves to be a successful strategy in forecasting cryptocurrency prices.

A significant finding from the Vo et al. (2019) study highlights the enhanced predictive power when sentimental data from news sources is combined with historical price information for cryptocurrencies. This approach exceeds the accuracy of using historical data alone. Specifically, the research demonstrated the effectiveness of utilizing news data (sourced from NewsNow) in forecasting price movements of Ethereum. The developed model in this study is proficient in determining the future direction of cryptocurrency prices, guiding decisions to buy, sell, or hold. A key insight from this study is the critical role of market psychology, which is often overlooked in trading scenarios. Additionally, the study emphasizes that sentiment analysis is a critical perspective for cryptocurrency price prediction due to the interactive nature of financial activities.

Certain studies have employed technical indicators as a key tool in analyzing and predicting market trends. In a study by McNally et al. (2018) exploring the viability of using machine learning techniques for predicting Bitcoin prices, a notable aspect was the use of technical analysis indicators, particularly simple moving averages (SMA), to enhance the predictive models. The SMA, a widely acknowledged tool for forecasting stock market trends, is among the various methods employed in time series analysis. It determines stock trends by averaging stock prices over a specific period(Lauren & Harlili, 2014). In the research conducted by Pichaiyuth et al. (2023), machine learning models enhanced with SHAP for feature selection demonstrate a notable proficiency in short-term cryptocurrency price trend forecasting. These models leverage a comprehensive suite of technical indicators, including Moving Average Convergence Divergence (MACD), Relative Strength Index (RSI), Stochastic RSI (StochRSI), Williams Percent Range (Williams %R), Price Rate of Change (ROC), Commodity Channel Index (CCI), Average Directional Index (ADX), and Simple Moving Average (SMA), all derived from daily high, low, and closing price data. The study finds distinct advantages in using different models for varying forecast periods, with particular effectiveness observed for longer 90-day forecasts. This underscores the critical role of SHAP in enhancing the predictive accuracy of financial trend analysis models, especially when integrated with an extensive array of price-based technical indicators.

3.2 *Market phases*

Next, in a 2020 study on bull and bear markets, Jankovic highlights two phases of extended intervals during which equity prices consistently either rise or fall. These distinct phases are recognized as: bull and bear markets.

However, a notable challenge arises from the lack of consensus among researchers and practitioners regarding how to precisely define the starting and ending dates of these bull and bear markets (Jankovic, n.d.). One simple method to detect market changes is through the use of Simple Moving Averages (SMA). Investors use SMA to assess both medium and long-term trends over an extended period. When a stock's 200-day SMA dips below its 50-day SMA, it is often viewed as a bear market. On the other hand, the opposite pattern suggests a possible bull market (Hayes, 2023). This method is sometimes utilized to identify the transitions between bull and bear markets in financial analysis (Edvall & Höjlind, 2020).

Cryptocurrency markets, much like traditional markets, undergo distinct phases of growth, stagnation, and correction. In a study by Suda and Spiteri (2019a), it was discovered that Bitcoin experiences more volatile bull and bear markets compared to the S&P500. In another study, Suda and Spiteri (2019b) revealed distinct volatility and persistence patterns in Bitcoin/USD compared to the S&P 500, with Bitcoin exhibiting high mean values and strong persistence in bull markets, and large negative mean values with weaker persistence in bear markets. This pivotal finding emphasizes the unique behavioral characteristics of the cryptocurrency market in comparison to traditional stock indices.

An interesting study by Kim et al. (2022) found that the cryptocurrency market is more responsive to positive social sentiment during downward trends and to negative sentiment during upward trends. This study used six months of data each from a bull and a bear market. Key features analyzed included Bitcoin trading volume and closing price, as well as social media sentiment indicators like Google trend rate, and positive and negative sentiment rates on Twitter. When the market experiences a local downward trend, it exhibits greater sensitivity to positive social sentiment. In contrast, during an upward local trend, the market is more influenced by negative social sentiment (Kim et al., 2022).

3.3 *Price prediction techniques*

In this subsection, the methodologies for forecasting the Bitcoin price are discussed. The exploring the world of time series forecasting, many studies use advanced methods to find patterns and insights. LSTM models are recognised in the literature as the most advanced, or state-of-the-art (SOTA), methods specifically for (sequence learning) forecasting Bitcoin prices, as referenced in Wu et al. (2018). The LSTM is an artificial RNN architecture frequently employed in deep learning. It is capable of integrating an entire dataset, going beyond the analysis of individual data points (Andi, 2021). Across numerous studies (Andi, 2021; Kumar & Rath, 2020; Latif et al., 2023; McNally et al., 2018; Patel et al., 2020; Pathak & Kakkar, 2020; Serafini et al., 2020; Vo et al., 2019; Wu et al., 2018) , the LSTM models were commonly employed, with a significant number of these investigations highlighting the LSTM as the superior model in terms of performance.

In the study by Kumar and Rath, the Multi-Layer Perceptron (MLP) and LSTM models were compared. In the comparative analysis, both MLP and LSTM models proved capable of predicting the price trends of Ethereum, yet LSTM displayed a more robust and precise capability for handling long-term dependencies compared to MLP. Wu et al. (2018) research developed two LSTM-based models. The first was a conventional LSTM . The second LSTM model used autoregressive (AR) components, notably AR(2). Results showed that the LSTM-AR(2) model had superior forecasting accuracy compared to the standard LSTM, highlighting the advantages of incorporating autoregressive elements into predictive models.

The study of Pichaiyuth et al. (2023) compared the effectiveness of SVM, KNN, RFC, Naïve Bayes, and LSTM in managing time-series data and recognizing patterns, finding that while SVM outperformed others in 7, 15, and 30-day forecasts, LSTM excelled in predictions over a 90-day period. The LSTM-GRU hybrid model, as proposed in the study by Patel et al., demonstrated enhanced accuracy over traditional LSTM and GRU models, as evidenced by consistently lower Mean Squared Error (MSE) values across various window sizes (1, 3, 7, and 30 days) for both Litecoin and Zcash.

In a study conducted by (Serafini et al., 2020), findings revealed that the LSTM model was not the most effective approach. In that study, a comparison between the Auto-Regressive Integrated Moving Average with eXogenous input (ARIMAX) and the LSTM model revealed that ARIMAX outperformed LSTM, suggesting that linear models can be highly effective in forecasting the Bitcoin market. Iqbal et al. (2021) demonstrated that the ARIMA algorithm surpassed more complex models, such as FBProp and XGBoost, in forecasting Bitcoin prices using historical data. Meanwhile, Pathak and Kakkar (2020) emphasized the prominence of the LSTM model in time series forecasting. The study of Inamdar et al. used a Recurrent Neural Network (RNN) with LSTM units for Sentiment Analysis, which effectively processed extensive sequences from Twitter and news. For Price Prediction, they employed the Random Forest Regression algorithm, utilizing historical data and sentiment scores.

3.4 *Feature importance analysis*

Methods of Explainable Artificial Intelligence (XAI) are commonly used to make the workings of complex machine learning models more transparent.

However, the application of XAI in time series is still relatively unexplored (Saluja et al., 2021). Saluja et al. (2021) show that model-agnostic explanation techniques, such as LIME and SHAP, can effectively demystify time series forecasting models. Additionally, a human evaluation of the study revealed that both LIME and SHAP were instrumental in elucidating the model predictions. The use of these methods is crucial for improving the clarity and dependability of machine learning models for users. This factor becomes particularly critical in business settings, where it is essential to understand how various factors influence outcomes. In another study, researchers enhanced the interpretability of a complex deep learning model used for air quality time series prediction. To interpret the predictions, the study applied the SHAP framework, which significantly clarified the behavior of the model, making the predictions more transparent (García & Aznarte, 2020).

In an interesting study by Carbó and Gorjón (2022), an LSTM neural network and SHAP interpretability techniques were used to analyze the changing influence of technological, economic, and investor attention variables on the Bitcoin price across various time periods. SHAP's insights reveal that while technological factors like hash rate and block size were key determinants in Bitcoin's price in earlier periods. However, their effectiveness has gradually decreased over time. In contrast, the importance of public attention variables, like Google Trends searches and Twitter mentions, has grown, becoming the most influential factors in later periods. This highlights a significant shift in the drivers of Bitcoin's value.

3.5 *Scientific gap*

The literature on Bitcoin price prediction underscores the importance of historical data as a robust indicator (Andi, 2021; Iqbal et al., 2021), while sentiment data from news and social media platforms influences cryptocurrency price dynamics (Abraham et al., 2018; Valencia et al., 2019; Wołk, 2020). In studies such as those by (Latif et al., 2023; Serafini et al., 2020; Wu et al., 2018), the predictions appear to be shifted along the x-axis. This is a common issue where a model erroneously predicts future values by simply mirroring past trends, rather than accurately forecasting based on current market dynamics. The study addresses this by not only pinpointing the most effective models for Bitcoin price forecasting but also rigorously examining their behavior to prevent 'time series shifting'. While Kim et al. (2022) discusses bull and bear markets and the role of sentiment, the literature lacks an in-depth exploration of which specific features are most influential during these distinct market phases. Notably, a detailed feature importance analysis to identify key influential factors in different market conditions is absent.

This study aims to fill this gap by providing a comprehensive and novel understanding of the effectiveness of historical prices and sentiment data in predicting Bitcoin prices across bull and bear markets. While the study by (Carbó & Gorjón, 2022) provided valuable findings using SHAP analysis, it focused on broader time phases rather than a direct comparison of bull and bear markets. This research proposes to extend this approach by specifically examining how influential factors diverge in their impact between bull and bear markets. This comparative analysis could offer a more granular understanding of market drivers in different economic climates, thereby enhancing the overall comprehension of market behavior.

4 METHOD

In this chapter, the core methodologies of this study are outlined. First, the baseline model for time series forecasting will be discussed. Then, the focus shifts to RNN with an emphasis on LSTM and GRU models. Following this, the approach for detecting Bull and Bear Market phases is described. Lastly, model interpretability will be emphasized to understand the drivers behind the model predictions.

4.1 *Baseline model*

The ARIMAX model has been selected as the baseline model for this study, given its frequent use in literature as a point of comparison with more complex models (Iqbal et al., 2021; Latif et al., 2023; McNally et al., 2018; Serafini et al., 2020). This model is an advanced extension of the ARIMA framework, designed to enhance time series forecasting by incorporating external variables. The structure of the model combines the autoregressive (AR) component, which models the dependency of the current value on its previous values. The integrated component (I) involves differencing the series to achieve stationarity. Differencing ensures consistent mean, variance, and autocorrelation over time. This makes the parameters of the model reliable for forecasting. Additionally, the moving average (MA) aspect accounts for the influence of past forecast errors on the current prediction (Nau, 2023). The distinctive feature of the ARIMAX model lies in its integration of exogenous variables (X), which allows for the inclusion of external influences that might affect the time series.

4.2 *Recurrent neural networks*

Recurrent Neural Networks, were first introduced by Rumelhart et al., 1986. They are a type of neural network designed specifically for handling sequential data. RNNs are adept at processing sequences of values, such as time series or text data. This specialization makes RNNs particularly useful in fields where understanding the context and order of data is crucial Goodfellow et al., 2016. In this study, two forms of RNNs, LSTM and GRU, will be used to forecast the Bitcoin price.

4.2.1 *Long Short-Term Memory*

The LSTM model was first introduced by Sepp Hochreiter and Jürgen Schmidhuber in 1997 Hochreiter and Schmidhuber, 1997. This model has been widely recognized for its effectiveness in handling complex time series data and sequences across various domains, including natural language processing, speech recognition, and financial forecasting. The LSTM model operates using three essential components: the input, output, and forget gates. The input gate controls how new information is incorporated. The forget gate, an important feature of this model, is key in deciding what past information to keep or discard over time. Lastly, the output gate selects what information from the model's internal memory should be used, considering both new inputs and what has been remembered so far.These gates collaboratively update the cell and hidden states at each time step, equipping the LSTM to effectively handle long sequences and address the vanishing gradient problem (Hochreiter, 1998).

4.2.2 *Gated Recurrent Unit*

The Gated Recurrent Unit (GRU) model was introduced by Cho et al. (2014). The GRU has been effectively utilized in the study conducted by (Patel et al., 2020), demonstrating its applicability and efficiency in modeling complex sequential data. This model simplifies the recurrent network structure by integrating two main gates: the update gate and the reset gate. The update gate is responsible for balancing the retention of previous information against the inclusion of new information, allowing the model to preserve long-term dependencies without the risk of vanishing gradients (Sherstinsky, 2020). The reset gate, conversely, determines how much past information to forget, thereby enabling the model to adapt to new data patterns as they emerge. The GRU model, in contrast to more complex models such as the LSTM, combines the forget and input gates into a single update gate, and merges the cell state and hidden state. This simplification leads to fewer parameters and a more efficient learning process, making the GRU a popular choice for tasks where computational efficiency is key.

4.3 *Bull and bear market detection*

In this study, the Simple Moving Average (SMA) will be employed as a key tool for detecting bull and bear markets. In the context of financial markets, the emergence of bear markets is typically heralded by the occurrence of a 'death cross' in the trading graphs of stocks or ETFs, where the 50-day SMA dips below the 200-day SMA Insider, 2020. Conversely, a bull market is often indicated by a 'golden cross', where the 50-day SMA rises above the 200-day SMA, signaling upward market momentum.

4.4 *Model interpretability*

In this thesis, the focus is on employing SHAP, a method developed by Lundberg and Lee (2017b), to elucidate individual predictions. SHAP is grounded in the game-theoretically optimal Shapley values, providing a solid theoretical foundation for explaining model predictions. Specifically, SHAP evaluates the importance of each feature by calculating its average marginal contribution to the model's prediction across all possible combinations of features, aggregating these effects to determine the feature's overall impact on the model's predictions. This method's application aims to offer clear insights into how different features impact the model's output. SHAP relies on the magnitude of feature contributions to assess their importance Molnar, 2017. The study aims to provide an understanding of feature contributions in the context of the model.

5 EXPERIMENTAL SETUP

In this chapter, the experimental setup employed to address the research questions is outlined. Figure 1 offers a comprehensive overview of this setup. Each step of the experiment, from data acquisition to final analysis, is detailed in the subsequent sections.

5.1 *Input data*

This study compiles a comprehensive dataset over 2,421 days (November 1, 2016, to June 18, 2023), incorporating both market prices and public sentiment, offering insights into market trends and sentiment fluctuations during this period.

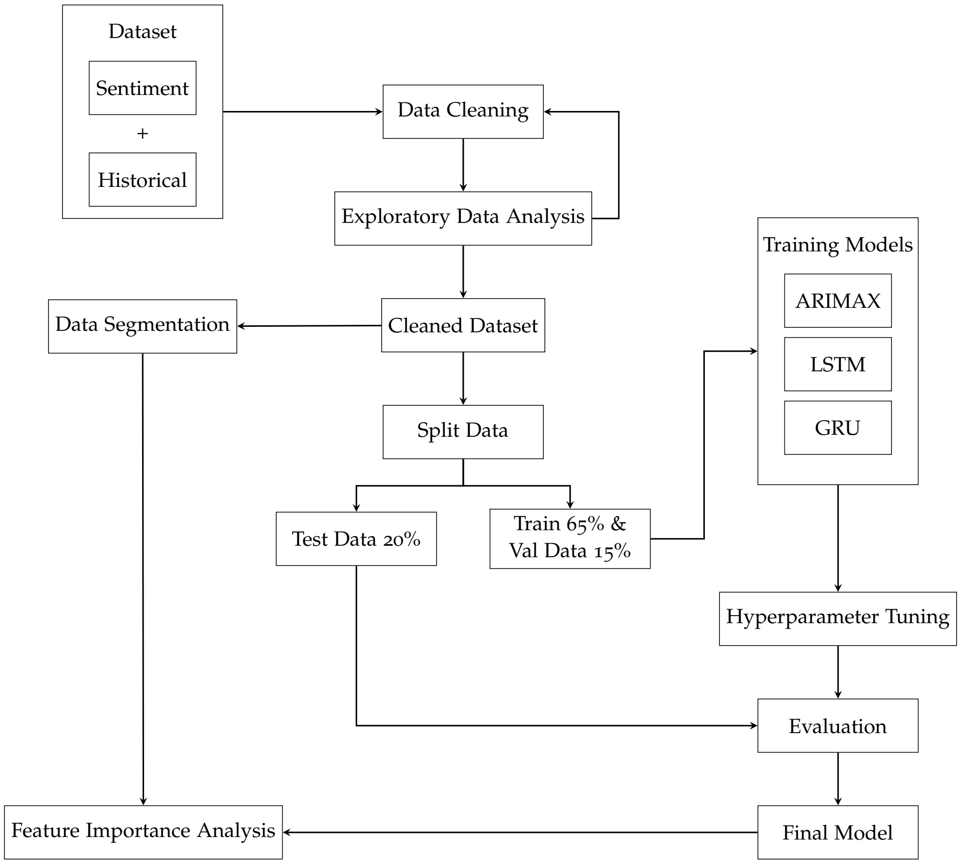

Figure 1: Experimental Setup Flowchart: Initiates with dataset acquisition, including sentiment and historical data, leading to data cleaning and EDA. Following this, the dataset undergoes segmentation for detailed analysis. Subsequently, it's divided into distinct sets: training/validation (65%/15%) and test (20%). The workflow includes training of ARIMAX, LSTM, and GRU models, hyperparameter tuning, and model evaluation. The final stage involves selecting the best model and conducting a focused feature importance analysis.

5.1.1 *Price data*

Historical price data are obtained from Yahoo Finance Yahoo Finance, 2023, covering the aforementioned period with 2,421 records. The dataset includes critical features such as Date, Open, High, Low, Close, Adjusted Close, and Volume, integral to the analysis.

5.1.2 *Sentiment data*

Sentiment data are sourced from Augmento.ai, which analyzes cryptocurrency-related discussions on platforms like Twitter (X), Reddit, and Bitcointalk. This dataset, available in CSV format, spans from 2016-11-01 01:00:00 to 2023-06-18 06:00:00 and comprises 58,086 entries across 95 features, including 93 distinct sentiment categories (Augmento, 2020). For this study, only the 14 features representing positive and negative sentiments are utilized, as detailed in Table 1. These datasets align with the study's objectives and support the research questions.

Table 1: Classification of Sentiment Features. Outlines positive and negative sentiment features used in the analysis.

Positive	**Negative**
Bullish	Bearish
Optimistic	Pessimistic/Doubtful
Happy	Sad
Euphoric/Excited	Fearful/Concerned
Positive	Angry
	Mistrustful
	Panicking
	Annoyed/Frustrated
	Negative

5.2 *Exploratory data analysis*

The primary goal of the Exploratory Data Analysis (EDA) chapter is to enhance understanding of the underlying patterns and relationships in the dataset. This involves examining distribution and boxplots, and conducting correlation analysis to identify relationships between variables. The chapter also delves into feature engineering to uncover more insights and optimize the dataset for further analysis. Additionally, the Augmented Dickey-Fuller test will be employed to check data stationarity, a crucial aspect for time-series analysis. Finally, the Granger causality test is applied to explore predictive relationships between series, further enriching the understanding of the dataset.

5.2.1 *Distribution and box plots*

The distribution plots revealed that the sentiment features, including twitter_sent_score, reddit_sent_score, bitcointalk_sent_score, and total_sent_score, are normally distributed. This normal distribution indicates a balanced representation of sentiment scores across the dataset. On the other hand, financial features such as open, high, low, close, and volume are notably right-skewed. Additionally, the boxplots highlighted the presence of some outliers, particularly in the financial features. However, these outliers have been retained in the analysis, considering the inherent high volatility of Bitcoin. In the context of cryptocurrency markets, extreme values are not uncommon and often represent significant market events or reactions. The detailed distribution and box plots are presented in Appendix A for reference.

5.2.2 *Correlation analysis*

The correlation matrix shown in Figure 2 reveals a multifaceted relationship between various financial metrics and sentiment scores from social media platforms. Firstly, the trading metrics of Bitcoin, including the open, high, low, and close prices, exhibit a remarkably high correlation with each other, with correlations of 1. This indicates a strong linear association among these variables, meaning that movements in one are closely mirrored by the others. Such a scenario is an example of multicollinearity, a phenomenon where two or more predictor variables in a statistical model are highly correlated. This high level of multicollinearity makes it challenging to isolate the individual effect of each predictor on the response variable, as noted in the study of (Allen, 1997). To address this issue and improve the clarity of the models, the 'open', 'low', and 'high' columns will be removed from the analysis, focusing instead on the 'close' prices and other less correlated variables. In addition to this, new features will be added, capturing essential aspects and trends within the data. The volume of trades shows a moderate positive correlation with the trading metrics, hovering around 0.62. This reflects a moderate linear association. In contrast, the correlation of these trading metrics with sentiment scores from platforms like Twitter, Reddit, and Bitcointalk is notably weaker. The sentiment scores, particularly from Twitter and Bitcointalk, show a slight negative correlation with the trading metrics, with coefficients ranging from -0.03 to -0.29. This suggests a minimal linear relationship.

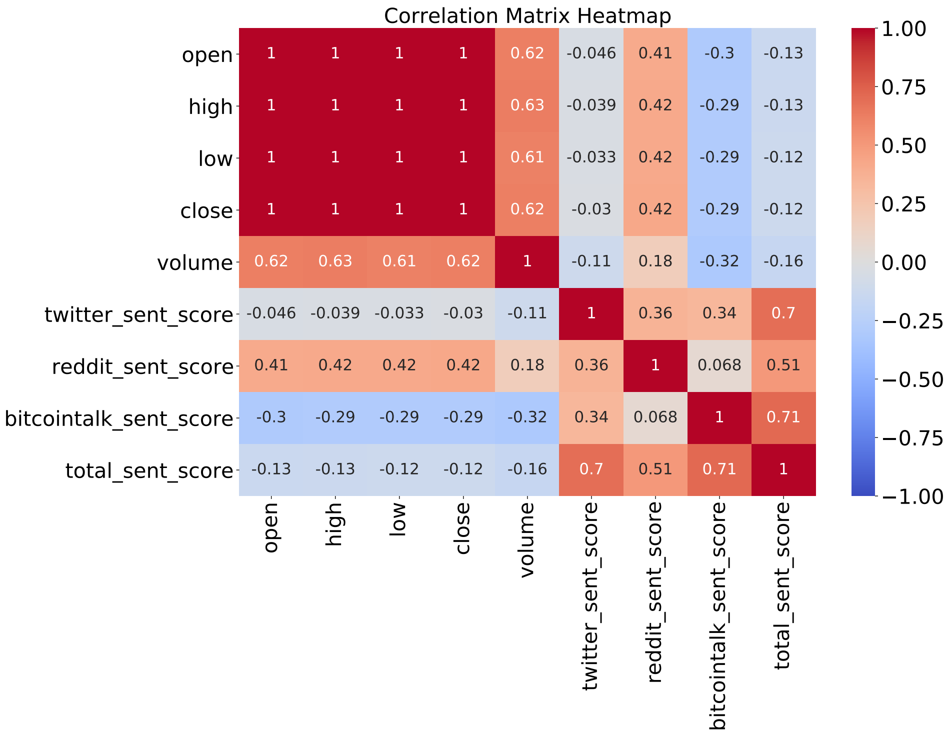

Figure 2: Correlation Matrix

5.2.3 *Feature engineering*

Feature engineering involves the process of transforming raw data into meaningful patterns, simplifying the task for machine learning models to make accurate predictions. It is often regarded as a crucial aspect of the data mining process, essential for achieving optimal outcomes in predictive modeling (Zheng & Casari, 2018).

The sentiment dataset requires some feature engineering to obtain a viable daily sentiment score. The features that are created are the sentiment scores for Twitter, Reddit, Bitcointalk, and a combined score, which will be determined using the following formula(AlphaSense, 2023):

$$\text{Sentiment Score} = \frac{\text{Total positive sentiment} - \text{Total negative sentiment}}{\text{Total positive sentiment} + \text{Total negative sentiment}}$$

This score will range from -1 to 1, with -1 indicating negative sentiment and 1 indicating positive sentiment.

Further expanding on this, as highlighted in Paragraph 5.2.2, the study introduces 'daily_range' and 'price_change' as new features. The 'daily_range' is derived from the difference between the highest and lowest prices of that day, offering insights into the extent of price fluctuation within a single day. On the other hand, 'price_change' is calculated as the difference between the closing and opening prices, providing a clear indication of the overall price movement for the day. These features are designed to capture essential aspects of market dynamics, focusing on key movements and trends, thereby enhancing the model's analytical capabilities.

To effectively capture the movements of the close price in the analysis, additional features, namely the Simple Moving Average (SMA) and the Exponential Moving Average (EMA), have been introduced. These features are only used for the deep learning models mentioned in Paragraph 4.2. The SMA is a type of arithmetic moving average calculated by averaging the closing prices over a specified time period (n). This method provides a smoothed representation of price trends over time. In contrast, the EMA, an advanced version of the moving average, applies more weight to recent data points, resulting in a more responsive indicator to recent price changes. This increased sensitivity makes the EMA a valuable tool for detecting short-term price movements and potential market reversals.

5.2.4 *Augmented Dickey-Fuller test*

For ARIMAX, ensuring the target variable is stationary is crucial. Stationarity means that the statistical properties of the process generating the time series data do not change over time. To verify if the target is stationary, the Augmented Dickey-Fuller test (ADF) was used. The ADF test checks for a unit root in a time series, which indicates non-stationarity. If a time series has a unit root, it shows a systematic pattern that could be evolving over time, making it unpredictable and unstable for modeling. The ADF test helps in determining if the series is suitable for ARIMA modeling by testing if it needs differencing to become stationary. This is important because ARIMA models require stationary data to make accurate and reliable forecasts. The target variable, the 'close' price, was initially non-stationary. To address this, it was converted into a stationary series by differencing the values.

5.2.5 *Granger causality test*

A Granger causality test was conducted to assess the potential of various time series features in forecasting the closing price data. This test specifically focused on determining whether historical sentiment time series could predict future values of price data. The sentiment scores from platforms like Twitter, Reddit, and Bitcointalk, as well as the aggregated total sentiment score, exhibited significant predictive power for the closing price data from the very first lag. In contrast to the sentiment scores, the 'price_change' feature did not show significant causality in most lags. This suggests that price changes may not be reliable predictors of future closing prices in this specific analysis. To comprehensively evaluate this predictive capability, Granger-Causality tests were performed across a range of lag values. The results, including the identification of the optimal number of lags for the model, are detailed in Appendix B, Table 7.

5.3 *Data preprocessing*

In this section, the key steps taken to prepare the dataset for analysis are outlined. The process begins with data cleaning, an essential phase for ensuring data quality and reliability. This is followed by data splitting, a crucial step for effective model training and accurate evaluation on unseen data. Subsequently, a normalization technique is explored, a vital step for uniformly scaling the data

5.3.1 *Data cleaning*

In the initial analysis of the datasets, it was found that missing values were present only in a redundant column, which was subsequently removed. However, after creating the sentiment features as detailed in Paragraph 5.2.3, a small number (7 instances) of missing values emerged. These were addressed by employing forward fill, propagating the last valid observation forward. This method was chosen over deletion to maintain continuity in dates, a necessity for the recurrent models used. Outliers, as analyzed in Paragraph 5.2.1, were retained due to Bitcoin's inherent volatility, reflecting typical market behaviors.

5.3.2 *Splitting the data*

For model optimization and evaluation, a hold-out method was applied, segmenting the dataset into 65% for training, 15% for validation, and 20% for testing. This split allowed for comprehensive model training while also providing sufficient data for tuning hyperparameters during validation. The 20% testing portion, used for out-of-sample evaluation, offered a reliable measure of the model's performance on unseen data, ensuring an accurate assessment of its predictive capabilities.

5.3.3 *Min-max normalization*

Throughout this study, min-max normalization has been implemented, as it aids in enhancing the performance of deep learning models Goodfellow et al., 2016. This approach helps in scaling the data uniformly, ensuring each feature contributes equally to the analysis. By using min-max normalization, the reliability and comparability of the results have been enhanced. To prevent data leakage, the scaling was based on the training data, and the same scale was applied to both the validation and test sets. This approach maintains the integrity of the evaluation by ensuring that the scaling remains consistent across training, validation, and testing datasets.

5.4 *Hyperparameters and features*

5.4.1 *Baseline model*

In the tuning of the ARIMAX model, key parameters are carefully adjusted, as detailed in Table 2. This tuning is important for optimizing the performance of the model.

Two sets of exogenous variables are tested to enhance the model's performance. The first set includes 'volume', 'daily_range', and 'price_change', focusing on market dynamics. The second set adds sentiment scores from Twitter, Reddit, Bitcointalk, and an overall sentiment score, introducing market sentiment aspects into the model.

Table 2: ARIMAX hyperparameter overview. Summarizes key model parameters and their value ranges.

Hyperparameters	**Description**	**Values**
AutoRegressive order (p)	Number of lag observations	1, 2, 3
Integrated order (d)	Number of times the observations are differenced	1
Moving Average order (q)	Size of the moving average window	1, 2, 3

5.4.2 *Deep learning models*

In the development of the LSTM and GRU models, specific set of parameters was utilized, which are detailed in Table3. The RNN input shape follows the [Samples, Timesteps, Features] format, with a chosen timestep of 10 for optimal results, while feature count varies by dataset. These parameters were carefully selected to optimize the performance of each model, ensuring that they are well-suited to the unique characteristics of the dataset and the objectives of the analysis. To prevent overfitting, a technique known as early stopping was employed. Early stopping is used to avoid overly complex models that perform well on training data but poorly on unseen data. By monitoring the performance on a validation set during training, early stopping halts the training process once the performance ceases to improve, thereby ensuring that the model remains generalizable to new data. The performance of the model was evaluated using the Mean Squared Error (MSE) as the metric. MSE is a common measure for regression models. The model that achieves the lowest MSE is considered the best performer, as it indicates the smallest average error in its predictions. The hyperparameter tuning process was conducted with the features as outlined in Table 4. These features were formulated to evaluate and compare the behavior of the model in terms of their predictions, with a specific focus on ensuring there is no shifting along the x-axis.

Table 3: LSTM and GRU hyperparameter overview.Summarizes key model parameters and their value ranges.

Hyperparameters	**Description**	**Values**
Nodes	Number of neurons in each layer	32, 64, 128
Dropout	Dropout for dropout layer	0, 0.1, 0.2
Batchsize	Number of training samples per iteration	32, 64, 128
Loss function	Metric used to measure the model's performance	MSE

Table 4: Feature Sets Overview. Outlines three sets of financial and sentiment analysis features.

Set	**Features**
Set 1	close, volume, daily_range, price_change *Including Sentiment:* twitter_sent_score, reddit_sent_score, bitcointalk_sent_score, total_sent_score
Set 2	close_diff, volume_diff, daily_range_diff, price_change *Including Sentiment:* twitter_sent_score, reddit_sent_score, bitcointalk_sent_score, total_sent_score
Set 3	close_diff, close_sma_10, close_sma_20, close_sma_100, close_ema_10, close_ema_20, close_ema_100, volume, daily_range, price_change *Including Sentiment:* twitter_sent_score, reddit_sent_score, bitcointalk_sent_score, total_sent_score

5.5 *Evaluation method*

5.5.1 *Evaluation metrics*

To assess the performance of the models, two key metrics are utilized: RMSE and MAE. RMSE measures the square root of the average squared differences between the predicted and actual values, providing insight into the model's accuracy by penalizing larger errors more severely. This makes RMSE particularly useful in highlighting significant errors in prediction. On the other hand, MAE calculates the average absolute difference between predicted and actual values, offering a straightforward interpretation of the average error magnitude. MAE is less sensitive to outliers compared to RMSE, making it a reliable measure of typical prediction errors. Together, RMSE and MAE give a comprehensive understanding of the model's precision and reliability in predictions, helping to identify areas for improvement.

5.5.2 *Loss plots*

In this study, loss plots are employed as a key tool for visualization. These plots are instrumental in demonstrating the degree of learning achieved by the model. Additionally, they provide clear indicators of when the model begins to overfit.

5.6 *Data segmentation*

The SMA approach, as outlined in Paragraph 4.3, was utilized to categorize the market into distinct phases. This involved calculating both the 50-day and 200-day SMAs, a method crucial for identifying the transitions between bull and bear markets. Each observation in the dataset was labeled as 'bull' or 'bear,' as shown in Figure 4.3. This labeling is based on the SMA calculations described in Paragraph 4.3, enabling a clear delineation of market conditions. This labeling process was instrumental in segmenting the entire dataset into smaller subsets, each representing a specific market phase. This segmentation into bull and bear periods allowed for a more targeted and nuanced analysis, facilitating a deeper understanding of market dynamics during different phases.

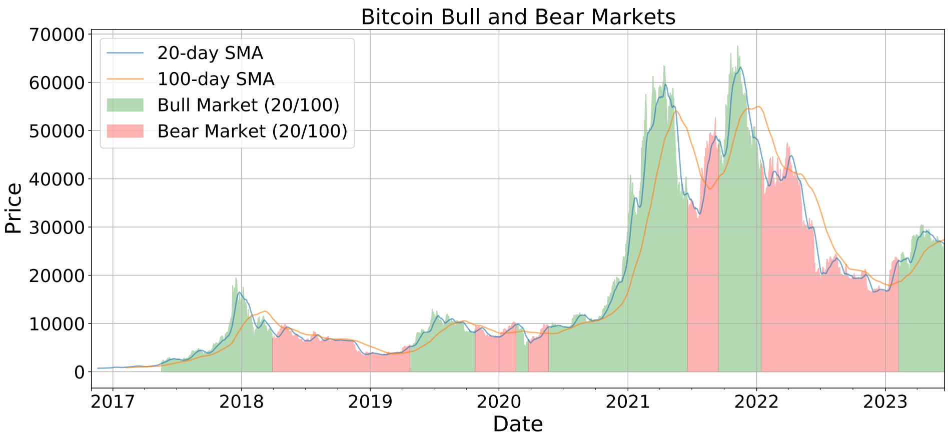

Figure 3: Bull and Bear Markets Segments

5.7 *Feature importance analysis*

Feature importance analysis is essential to comprehend the influence of individual features and feature groups on Bitcoin's behavior across bull and bear market phases. To accomplish this, SHAP was utilized to ascertain both individual and aggregated feature importance, shedding light on how these features distinctly affect Bitcoin's market dynamics during varying market conditions. SHAP is carefully applied to the results of models trained on the training data of each specific market segment that has been identified. The purpose of this train-test split is to ensure that the model learns to identify feature importance based on training data and then applies this learning to new, unseen testing data. This approach is critical for evaluating the model's ability to generalize and accurately identify feature importance in real-world scenarios. Furthermore, to account for the varying lengths of the bull and bear market segments, a weighting mechanism was introduced to the feature importance scores derived from SHAP. By multiplying the outcome of each feature importance score with a specific weight corresponding to the segment length, the influence of each segment is normalized, ensuring that longer segments do not disproportionately affect the overall importance analysis. The results of this weighted feature importance scoring are then aggregated, summing up the scores separately for bull and bear market conditions. This step offers a comprehensive understanding of the importance of individual features and grouped features across various market phases.

5.8 *Software used*

In this study, Python 3.8.5 was utilized, with key libraries including NumPy (Harris et al., 2020) for numerical operations, Pandas (McKinney, 2010) for data handling, and Matplotlib (Hunter, 2007) alongside Seaborn (Waskom, 2021) for visualization. Essential for scientific computations was SciPy (Virtanen et al., 2020). For machine learning, Scikit-Learn (Pedregosa et al., 2011), TensorFlow (Martín Abadi et al., 2015), Keras (Chollet, 2015), Statsmodels (Seabold & Perktold, 2010), and SHAP (Lundberg & Lee, 2017a) were employed. These libraries were instrumental in the data processing, analysis, and modeling conducted.

6 RESULTS

This section describes the results derived from the comprehensive analysis conducted using various modeling approaches: the baseline model, LSTM, GRU, and an extensive feature importance analysis.

6.1 *Model results*

In the model performance analysis, the ARIMAX model was initially established as the baseline, using volume, daily_range, and price_change as exogenous features. This model yielded an RMSE of 17,245 and a MAE of 15,934. The prediction of this model on the test data is shown in Figure 4, setting a preliminary standard for comparison with more complex models.

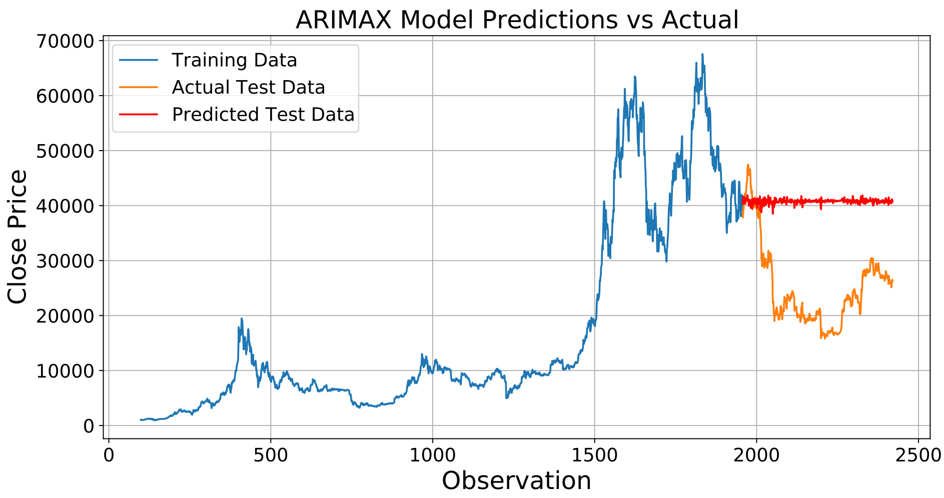

Figure 4: ARIMAX prediction with financial data

The more complex models, including the 'close' target as feature (set 1), shows the most favorable results. This indicates the significant predictive power of past 'close' prices in forecasting future values. However, as visualized in Figure 5, there seems to be a probable shift in the predictions of the model, suggesting that the model might be depending too much on the close feature to some extent.

To further explore the predictive capabilities, experimentation was conducted using the differenced closing price in the model (set 2) The results, illustrated in Figure 6 and Figure 7, showed that this approach revealed some predictive capabilities. In Figure 6, the differenced predictions made by the LSTM model are displayed. Subsequently, these results are converted back to their original scale by reversing the differencing process, as shown in Figure 7. In terms of model performance using only financial data, the GRU model with SMA and EMA features (set 3) emerged as the top performer, performing best with the least amount of shifting. As summarized in Table 5, this model achieved an RMSE of 1,408 and an MAE of 1,079, outperforming other models and feature sets. For comparison, the LSTM model with the same feature set obtained an RMSE of 2,895 and an MAE of 2,215. The predictions made by this optimal GRU model (set 3)

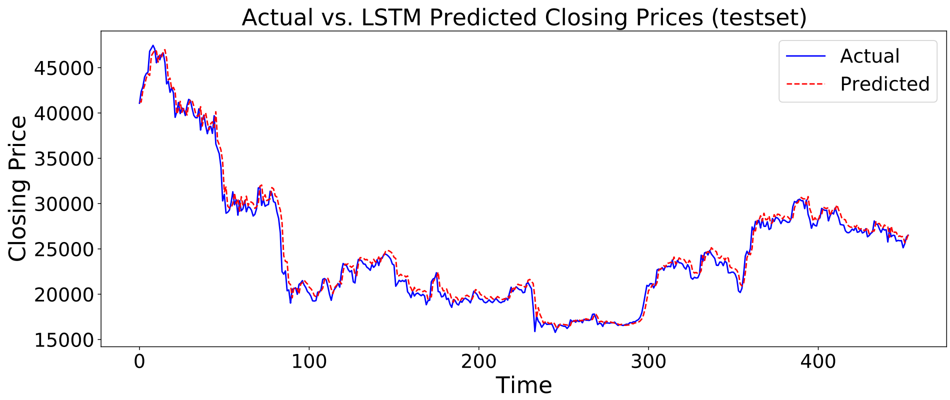

Figure 5: The prediction of the LSTM using 'close' feature only using financial data

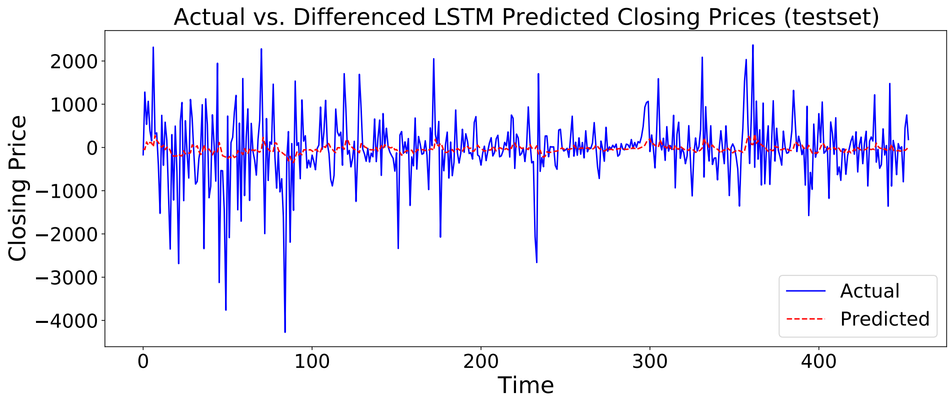

Figure 6: LSTM prediction with differenced financial data

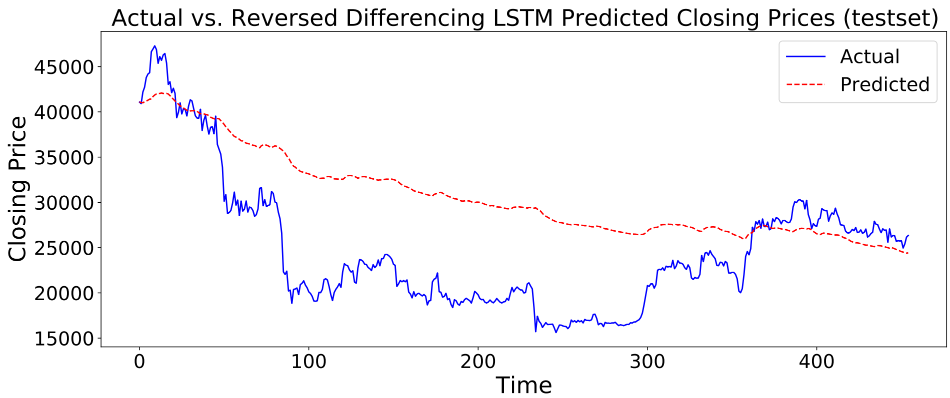

Figure 7: LSTM prediction with reversed differenced predictions

can be seen in Figure 8. This model was configured with the following hyperparameters: 64 nodes, a batch size of 32, 1 dense layer, the 'adam'

Table 5: Comparative Performance of ARIMAX, LSTM, and GRU Models Using Only Financial Data. This table presents the RMSE and MAE metrics for each model across different feature sets on the test dataset.

Model	Feature set	RMSE	MAE
ARIMAX	1	17,245	15,934
LSTM	1	935	664
	2	8,008	6,752
	3	2,895	2,215
GRU	1	989	719
	2	13,947	13,033
	3	1,408	1,079

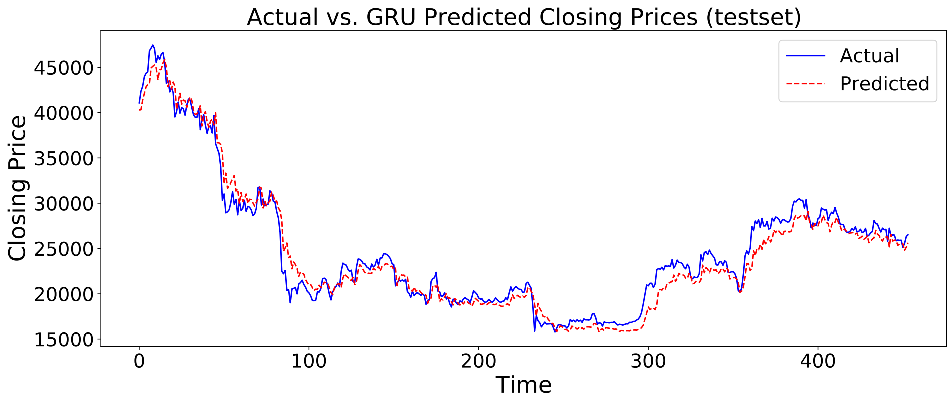

Figure 8: GRU model using SMA and EMA features

optimizer, and a dropout rate of 0.2. Additionally, the loss plot for this model is shown in Figure 9, illustrating the training and validation loss over epochs, further demonstrates the learning efficiency and stability of the model.

After incorporating sentiment features into the analysis, the models were re-evaluated to assess their predictive performance with this additional data. Surprisingly, the added sentiment features led to a decline in predictive performance, as detailed in Table 6. This outcome suggests that while sentiment data is valuable in understanding market dynamics, its direct correlation with Bitcoin price movements may not be as straightforward, or it might introduce additional noise into the predictive models.

Among the models tested with both financial and sentiment data, the models with the 'close' price as input feature (set 1) again stood out as the best performers. This result mirrored the similar shifting observed in

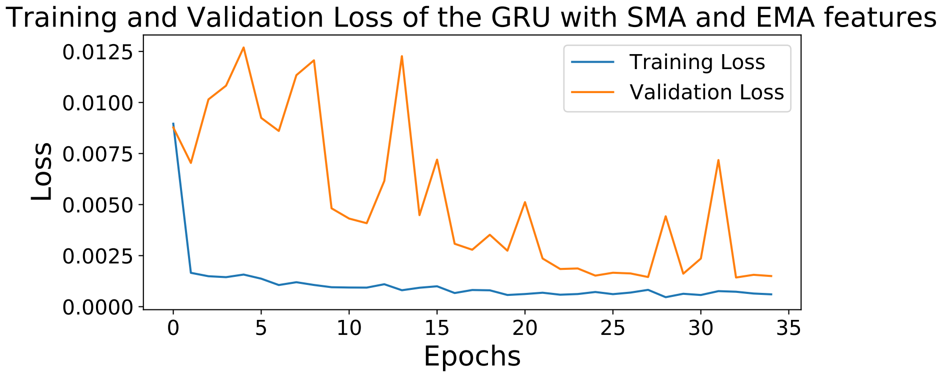

Figure 9: Loss of the GRU model using SMA and EMA features

the models without sentiment data, indicating a consistent pattern in the behavior of the model with the inclusion of the 'close' price feature.

The GRU model achieved an RMSE of 1,745 and an MAE of 1,248 with feature set 3, outperforming the other models and feature sets in this new context shown in Figure 10. This model configuration included 64 neurons, a batch size of 32, 1 dense layer, the 'adam' optimizer, and a dropout rate of 0.1. The effectiveness of these settings is further illustrated in Figure 11, which shows the loss plot for the model. For comparison, the LSTM model with the same feature set recorded an RMSE of 2,732 and an MAE of 1,866. The ARIMAX model, even with the additional sentiment features, recorded relatively high error metrics: an RMSE of 17,263 and an MAE of 15,948.

Table 6: Performance Metrics of ARIMAX, LSTM, and GRU Models on Test Data Incorporating Financial and Sentiment Data. The table compares RMSE and MAE values for each model variant across distinct feature sets.

Model	Feature set	RMSE	MAE
ARIMAX	1	17,263	15,948
LSTM	1	1,333	1,002
	2	56,783	48,411
	3	2,732	1,866
GRU	1	1,399	1,094
	2	86,725	75,589
	3	1,745	1,248

These results indicate that while the GRU model was most adept at handling the complexity of combined financial and sentiment data, the overall predictive performance across all models was negatively impacted by the addition of sentiment features. This finding highlights the challenges in integrating diverse data types and the importance of feature selection and model tuning in achieving optimal predictive performance in complex domains like cryptocurrency markets.

6.2 *Feature importance results*

The feature importance analysis is conducted using the best-performing model, namely the GRU network as shown in Paragraph 6.1. It is important to note that not all bull market segments could be utilized for this analysis. This limitation arose because some segments were too small to effectively train and test the model. Analyzing the SHAP values grouped by feature group for both 'bull' and 'bear' market conditions reveals insightful trends about the impact of financial and sentiment features on the models' predictions. The results, shown in Figure 12, clearly indicate that financial features are more influential than sentiment features in both market scenarios. Specifically, in bear markets, the 'Financial' feature group exhibits a notably higher SHAP value (0.019716) compared to that in bull markets (0.001793), suggesting that financial factors play a significant role

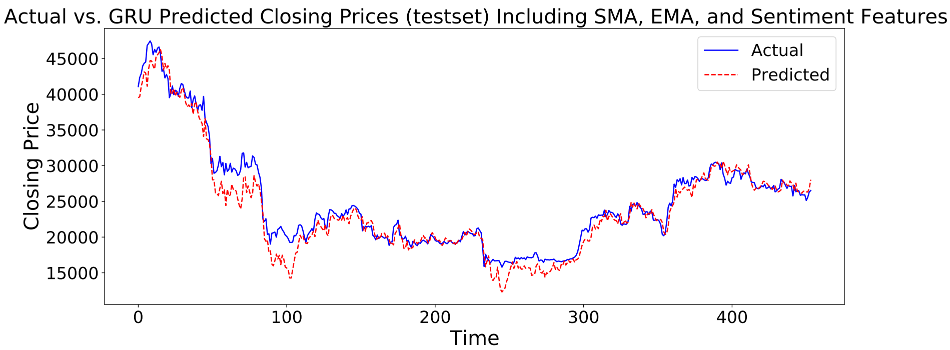

Figure 10: GRU model using SMA, EMA and sentiment features

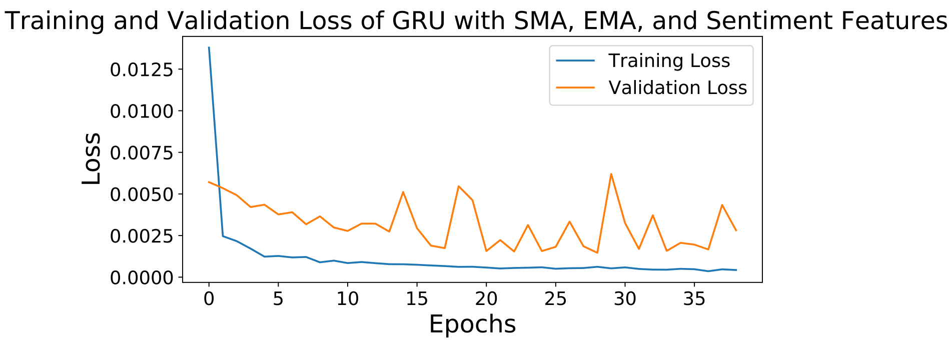

Figure 11: Loss of the GRU model using SMA, EMA and sentiment features

in driving predictions during downturns. This implies that in bearish conditions, traditional financial indicators may become crucial for accurate forecasting.

Conversely, sentiment features, while consistently contributing to the models' decision-making, have relatively lower SHAP values in both bull (0.001379) and bear (0.000018) markets. This trend indicates that while sentiment data provides valuable insights, it is secondary to financial data in terms of influencing the predictions of the model.

Delving deeper into the specific feature importances during these market phases, the results reveal distinct trends, as shown in 13. In bull markets, sentiment features such as twitter_sent_score, reddit_sent_score, and bitcointalk_sent_score exhibit notable SHAP values. For instance, the twitter_sent_score and reddit_sent_score demonstrate considerable influence with SHAP values of 0.001365 and 0.000977, respectively. This indicates a significant role of public sentiment in influencing the model's predictions during bullish conditions, where market optimism might be more sensitive to social media sentiment.

However, in bear markets, while sentiment features still contribute to the model's predictions, their relative importance diminishes compared to financial features. For example, the SHAP values for twitter_sent_score and bitcointalk_sent_score in bear markets are 0.000671 and 0.001089, respectively, which, although significant, are lower compared to their influence in bull markets. This trend suggests that during bearish conditions, where the market is typically characterized by uncertainty and negative sentiment, the model places greater emphasis on financial indicators like close_ema_10, which shows a markedly high SHAP value of 0.006907.

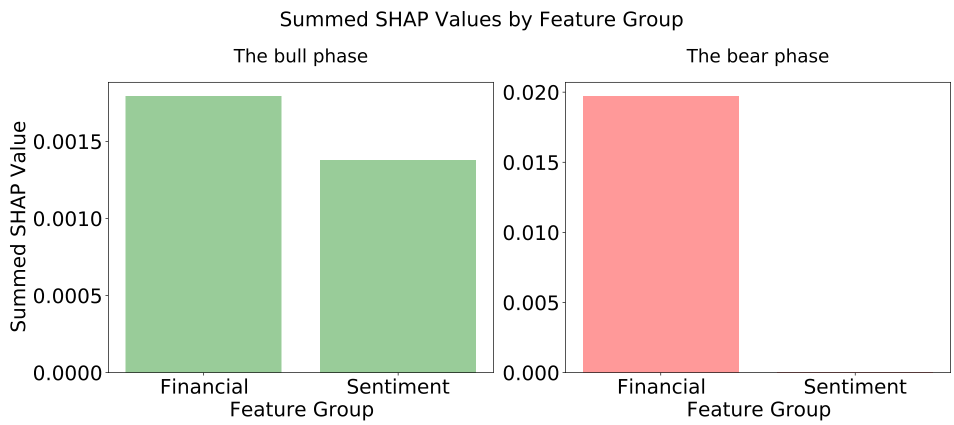

Figure 12: Feature Importance Per Feature Group

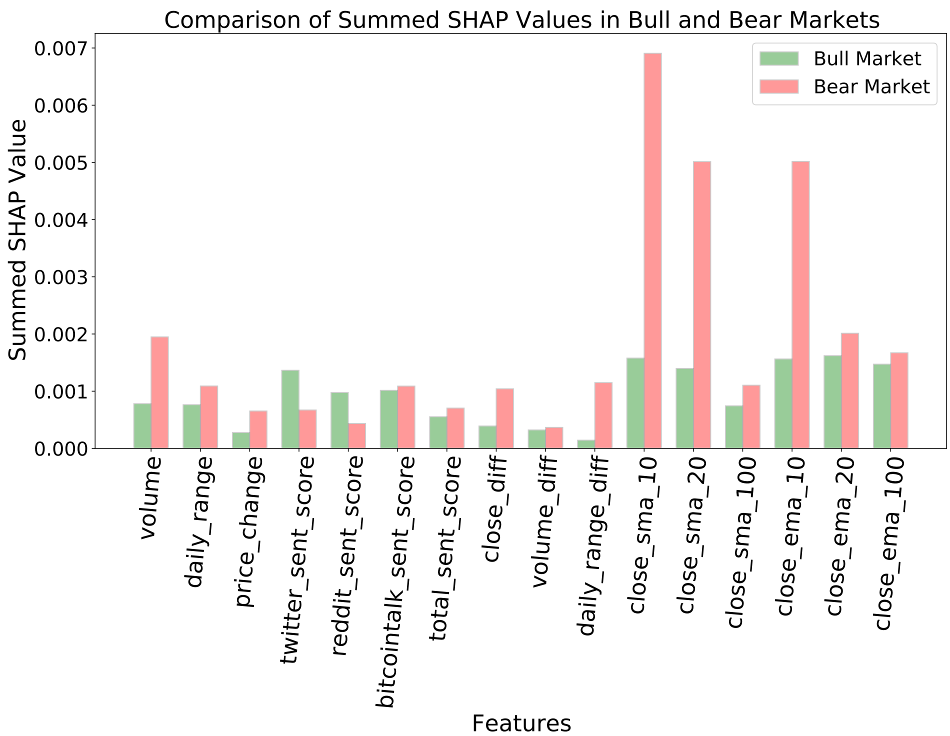

Figure 13: Feature Importance Per Feature

7 DISCUSSION

The purpose of this study was to investigate the impact of bull and bear market conditions on the effectiveness of historical data and sentiment data in forecasting Bitcoin prices. This exploration aimed to address the **Main Research Question:** To what extent do bull and bear market conditions impact the effectiveness of historical and sentiment data in forecasting Bitcoin price?

SQ1. *How effectively can the Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) models predict Bitcoin's value based solely on historical data compared to the Autoregressive Integrated Moving Average with Exogenous Variables (ARIMAX) model?* In this study, the LSTM and GRU models, particularly when utilizing the 'close' price as a feature, demonstrated superior effectiveness in predicting Bitcoin's value compared to the ARIMAX model. This superiority is quantitatively supported by significantly lower RMSE and MAE values. However, it was observed that these models exhibited a consistent shift in their predictions, suggesting they might be overly dependent on the 'close' price feature. This pattern was observed in models both with and without sentiment data. Interestingly, the study found that indirect use of the closing price, as incorporated in SMA and EMA features, resulted in more balanced and accurate predictions, thus mitigating the overreliance on direct 'close' price data.

SQ2 *To what extent do supplementary sentiment features influence the Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) of the LSTM and GRU models compared to the ARIMAX model?* When sentiment features were incorporated into the LSTM and GRU models, a noticeable decline in predictive performance was observed, as reflected by increased RMSE and MAE values. Although the GRU model (set 3) emerged as the best performer with the inclusion of both financial and sentiment data, its effectiveness diminished compared to its performance with purely financial data. This indicates that while sentiment features provide valuable market context, their integration into predictive models for Bitcoin's price forecasting may complicate model performance rather than enhance it, underscoring the challenges of effectively incorporating diverse data types into predictive financial models.

SQ3. *Which features emerge as the most influential when conducting a SHAP feature importance analysis on the best performing model during bull and bear market phases?* The SHAP feature importance analysis conducted on the GRU model reveals a distinct pattern in the influence of financial and sentiment features across bull and bear market phases as shown in Figures 12 and 13. Financial features consistently demonstrate a greater impact on the predictions of the model in both market scenarios. In bear markets, the prominence of financial features is significantly heightened, emphasizing their crucial role in forecasting during periods of downturn and uncertainty. While sentiment features also contribute to the predictions, their relative impact is less significant compared to financial features, though they hold notable importance in bull markets.

In the evolving landscape of Bitcoin price prediction, deep learning models have emerged as potent tools. Literature, notably through studies like (Andi, 2021; Patel et al., 2020), has underscored the effectiveness of LSTM models in leveraging historical data for accurate forecasting. However, this study reveals that the GRU model, especially when utilizing financial indicators such as SMA and EMA, performs better than other models. This finding aligns with the literature's focus on advanced models like LSTM and GRU, but it distinctly highlights the superior performance of the GRU model. This research thus contributes to the broader narrative by suggesting that while LSTM models are highly effective, GRU models might offer an enhanced predictive capability.

The role of sentiment data, as derived from social media and news sources, has been a topic of considerable interest in recent literature. Studies such as those by (Valencia et al., 2019; Wołk, 2020) have underscored the significant impact that public sentiment, as captured through these mediums, can have on the fluctuations of cryptocurrency prices. These findings suggest that the collective mood and opinions expressed online can provide valuable insights into market trends, thereby serving as a powerful predictor in forecasting models. However, in contrast to this view in the literature, the results of this study present a more complex picture regarding the integration and effectiveness of sentiment data. Upon incorporating sentiment features into the predictive models, a noticeable decline in performance was observed. This outcome suggests that while sentiment data holds potential value, its integration into predictive models for cryptocurrencies like Bitcoin is not straightforward and may introduce unforeseen complexities. This divergence between the general positive outlook on sentiment data in the literature and the findings of this study highlights the need for a deeper exploration of how sentiment data is processed, interpreted, and utilized within predictive models, ensuring that its integration does not compromise the accuracy and reliability of the predictions.

In the pursuit of advancing cryptocurrency price prediction research, future studies can aim to expand the application of SHAP analysis for a more detailed understanding of how various features influence price predictions across different market phases. Additionally, incorporating analysis of other major cryptocurrencies, such as Ethereum, can provide a comparative perspective and enrich the understanding of the broader cryptocurrency market. Exploring the integration of on-chain data and indicators reflective of significant news events could also offer a more comprehensive view of the factors driving cryptocurrency prices. This holistic approach would not only enhance the predictive models but also contribute to a deeper and more nuanced understanding of cryptocurrency market dynamics.

8 CONCLUSION

The primary aim of this study was to comprehensively investigate the extent to which the price of Bitcoin can be forecasted. A significant part of this investigation involved examining the potential of various sentiment features, sourced from diverse platforms, in adding value to the predictive models. These sentiment indicators underwent careful analysis to determine their impact on enhancing the accuracy and reliability of Bitcoin price predictions. However, the focal point of the study was to understand the behavior of these features across different market phases. This crucial aspect of the research illuminated how the influence of both price-related and sentiment-related features varied across distinct market conditions, including bull and bear phases. The findings from this analysis provided valuable insights into the complex dynamics of the cryptocurrency market, revealing relationships between market sentiment, price fluctuations, and market phases

In conclusion, this study makes a notable contribution to the field of cryptocurrency market analysis by shedding light on the varying importance of different features during distinct market phases. It underscores the complexities and challenges involved in effectively integrating sentiment data into predictive models for Bitcoin price forecasting. The findings suggest that while sentiment data from social media platforms offers a unique perspective on market perceptions, traditional financial metrics, particularly historical financial data, play a more dominant role in forecasting Bitcoin prices. This research underscores the need for a more sophisticated and nuanced approach to model construction, one that effectively balances and integrates the strengths of both traditional financial indicators and novel sentiment-based data.

REFERENCES

Abraham, J., Higdon, D., Nelson, J., & Ibarra, J. (2018). Cryptocurrency price prediction using tweet volumes and sentiment analysis. *SMU Data Science Review*, *1*(3), 1.

Allen, M. P. (1997). The problem of multicollinearity. *Understanding regression analysis*, 176–180.

AlphaSense. (2023). *Sentiment score: What it is and how to calculate it*. https://www.alpha-sense.com/blog/engineering/sentiment-score/

Andi, H. K. (2021). An accurate bitcoin price prediction using logistic regression with lstm machine learning model. *Journal of Soft Computing Paradigm*, *3*(3), 205–217.

Ari, L., & MacKenzie, S. (2022). *Crypto peaked a year ago — investors have lost more than* 2*trillionsince.CNBC*. https://www.cnbc.com/2022/11/11/crypto-peaked-in-nov-2021-investors-lost-more-than-2-trillion-since.html

Augmento. (2020). *Bitcoin sentiment – bull & bear index* [Accessed: yyyy-mm-dd]. https://www.augmento.ai/bitcoin-sentiment/

Carbó, J. M., & Gorjón, S. (2022). Application of machine learning models and interpretability techniques to identify the determinants of the price of bitcoin.

Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

Chollet, F. (2015). Keras.

Edvall, L., & Höjlind, J. (2020). Bear vs bull market: The difference in market behavior between the two phases.

García, M. V., & Aznarte, J. L. (2020). Shapley additive explanations for no2 forecasting. *Ecological Informatics*, *56*, 101039.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning* [http://www.deeplearningbook.org]. MIT Press.

Guardian, T. (2022). *Biden signs cryptocurrency order to examine risks as popularity rises*. https://www.theguardian.com/technology/2022/mar/09/biden-signs-order-cryptocurrency-amid-boom-popularity

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., . . . Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/s41586-020-2649-2

Hayes, A. (2023). *Simple moving average (sma): What it is and the formula*. https://www.investopedia.com/terms/s/sma.asp

Hochreiter, S. (1998). The vanishing gradient problem during learning recurrent neural nets and problem solutions. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, *6*(02), 107–116.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Hunter, J. D. (2007). Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, *9*(3), 90–95. https://doi.org/10.1109/MCSE.2007.55

Inamdar, A., Bhagtani, A., Bhatt, S., & Shetty, P. M. (2019). Predicting cryptocurrency value using sentiment analysis. *2019 International Conference on Intelligent Computing and Control Systems (ICCS)*, 932–934.

Insider, B. (2020). What bear market? the dow is officially in a bull market, and the s&p 500 may follow suit.

Iqbal, M., Iqbal, M., Jaskani, F., Iqbal, K., & Hassan, A. (2021). Time-series prediction of cryptocurrency market using machine learning techniques. *EAI Endorsed Transactions on Creative Technologies*, *8*(28).

Jankovic, M. (n.d.). Testing different garch models for bull and bear markets.

Kim, K., Lee, S.-Y. T., & Assar, S. (2022). The dynamics of cryptocurrency market behavior: Sentiment analysis using markov chains. *Industrial Management & Data Systems*, *122*(2), 365–395.

Kumar, D., & Rath, S. (2020). Predicting the trends of price for ethereum using deep learning techniques. *Artificial Intelligence and Evolutionary Computations in Engineering Systems*, 103–114.

Latif, N., Selvam, J. D., Kapse, M., Sharma, V., & Mahajan, V. (2023). Comparative performance of lstm and arima for the short-term prediction of bitcoin prices. *Australasian Accounting, Business and Finance Journal*, *17*(1), 256–276.

Lauren, S., & Harlili, S. D. (2014). Stock trend prediction using simple moving average supported by news classification. *2014 International Conference of Advanced Informatics: Concept, Theory and Application (ICAICTA)*, 135–139. https://doi.org/10.1109/ICAICTA.2014.7005929

Lundberg, S. M., & Lee, S.-I. (2017a). A unified approach to interpreting model predictions. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30* (pp. 4765–4774). Curran Associates, Inc. http://papers.nips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf

Lundberg, S. M., & Lee, S.-I. (2017b). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.

Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Jia, Y., Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, . . . Xiaoqiang Zheng. (2015). TensorFlow: Large-scale machine learning on heterogeneous systems [Software available from tensorflow.org]. https://www.tensorflow.org/

McKinney, W. (2010). Data Structures for Statistical Computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 56–61). https://doi.org/10.25080/Majora-92bf1922-00a

McNally, S., Roche, J., & Caton, S. (2018). Predicting the price of bitcoin using machine learning. *2018 26th euromicro international conference on parallel, distributed and network-based processing (PDP)*, 339–343.

Molnar, C. (2017). Interpretable machine learning. https://christophm.github.io/interpretable-ml-book/shap.html#fn69

Nau, R. (2023). Notes on nonseasonal arima models [Accessed: 2023-12-03].

NOS. (2022). *Les over cryptomunten moet mbo'ers wapenen tegen verleiding snelle geld*. https://nos.nl/artikel/2423184-les-over-cryptomunten-moet-mbo-ers-wapenen-tegen-verleiding-snelle-geld

Patel, M. M., Tanwar, S., Gupta, R., & Kumar, N. (2020). A deep learning-based cryptocurrency price prediction scheme for financial institutions. *Journal of information security and applications*, *55*, 102583.

Pathak, S., & Kakkar, A. (2020). Cryptocurrency price prediction based on historical data and social media sentiment analysis. *Innovations in Computer Science and Engineering: Proceedings of 7th ICICSE*, 47–55.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Pichaiyuth, P., Termnuphan, P., Triyason, T., Rojanapornpun, O., & Jaiyen, S. (2023). Price trend forecasting of cryptocurrency using multiple technical indicators and shap. *2023 20th International Joint Conference on Computer Science and Software Engineering (JCSSE)*, 150–154.

Pintelas, E., Livieris, I., Stavroyiannis, S., Kotsilieris, T., & Pintelas, P. (2020). Fundamental research questions and proposals on predicting cryptocurrency prices using dnns.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, *323*(6088), 533–536.

Saluja, R., Malhi, A., Knapič, S., Främling, K., & Cavdar, C. (2021). Towards a rigorous evaluation of explainability for multivariate time series. *arXiv preprint arXiv:2104.04075*.

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with python. *9th Python in Science Conference*.

Serafini, G., Yi, P., Zhang, Q., Brambilla, M., Wang, J., Hu, Y., & Li, B. (2020). Sentiment-driven price prediction of the bitcoin based on statistical and deep learning approaches. *2020 International Joint Conference on Neural Networks (IJCNN)*, 1–8.

Sherstinsky, A. (2020). Fundamentals of recurrent neural network (rnn) and long short-term memory (lstm) network. *Physica D: Nonlinear Phenomena*, *404*, 132306.

Suda, D., & Spiteri, L. (2019a). Analysis and comparison of bitcoin and s and p 500 market features using hmms and hsmms. *Information*, *10*(10), 322.

Suda, D., & Spiteri, L. (2019b). Comparing market phase features for cryptocurrency and benchmark stock index using hmm and hsmm filtering. *Business Information Systems Workshops: BIS 2019 International Workshops, Seville, Spain, June 26–28, 2019, Revised Papers 22*, 195–207.

Valencia, F., Gómez-Espinosa, A., & Valdés-Aguirre, B. (2019). Price movement prediction of cryptocurrencies using sentiment analysis and machine learning. *Entropy*, *21*(6), 589.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., . . . SciPy 1.0 Contributors. (2020). SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, *17*, 261–272. https://doi.org/10.1038/s41592-019-0686-2

Vo, A.-D., Nguyen, Q.-P., & Ock, C.-Y. (2019). Sentiment analysis of news for effective cryptocurrency price prediction. *International Journal of Knowledge Engineering*, *5*(2), 47–52.

Waskom, M. L. (2021). Seaborn: Statistical data visualization. *Journal of Open Source Software*, *6*(60), 3021. https://doi.org/10.21105/joss.03021

Wołk, K. (2020). Advanced social media sentiment analysis for short-term cryptocurrency price prediction. *Expert Systems*, *37*(2), e12493.

Wu, C.-H., Lu, C.-C., Ma, Y.-F., & Lu, R.-S. (2018). A new forecasting framework for bitcoin price with lstm. *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*, 168–175.

Yahoo Finance. (2023). Bitcoin usd (btc-usd) historical data.

Zheng, A., & Casari, A. (2018). *Feature engineering for machine learning: Principles and techniques for data scientists*. " O'Reilly Media, Inc.".

APPENDIX A

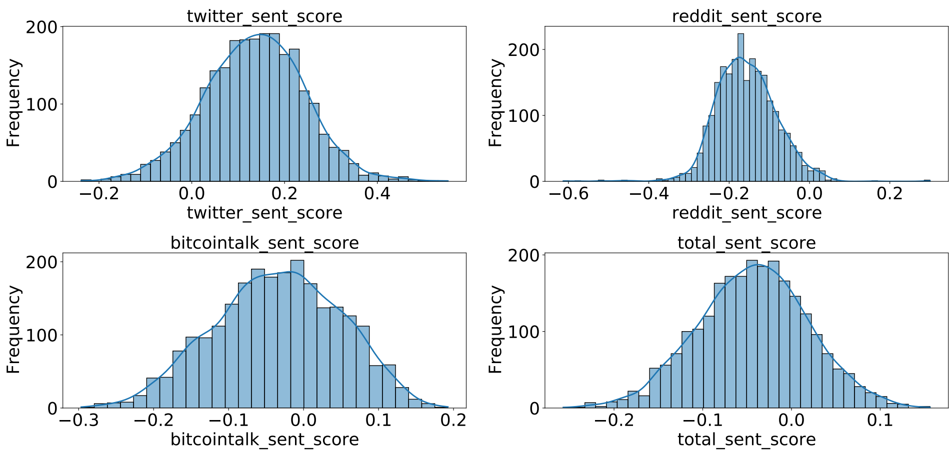

Figure 14: Histogram plots of the sentiment features

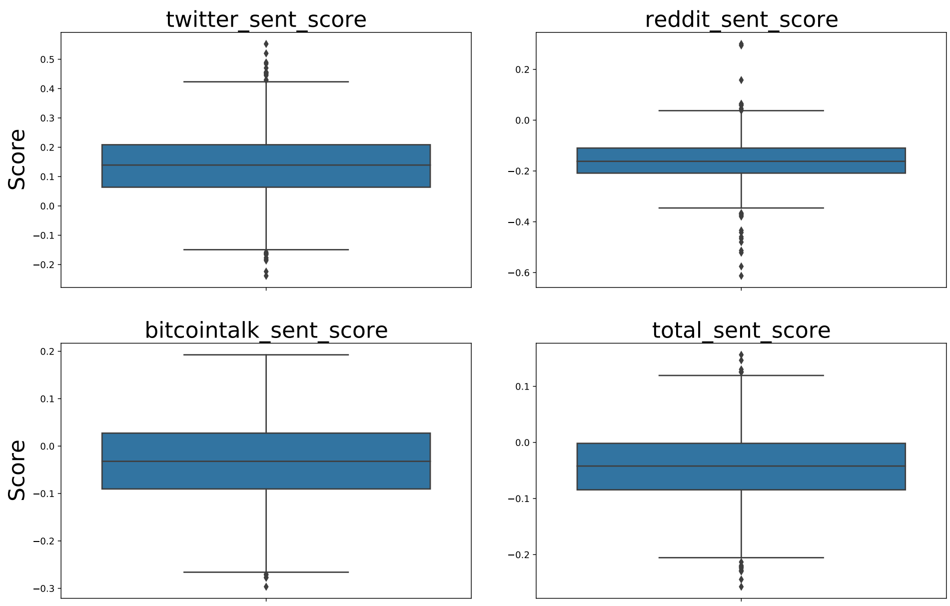

Figure 15: Box plots of the sentiment features

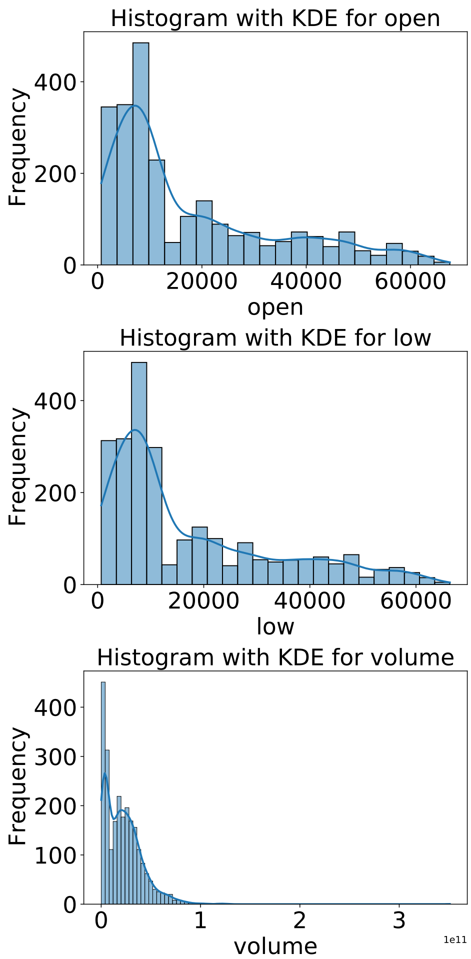

Figure 16: Histogram plots of the historical price features

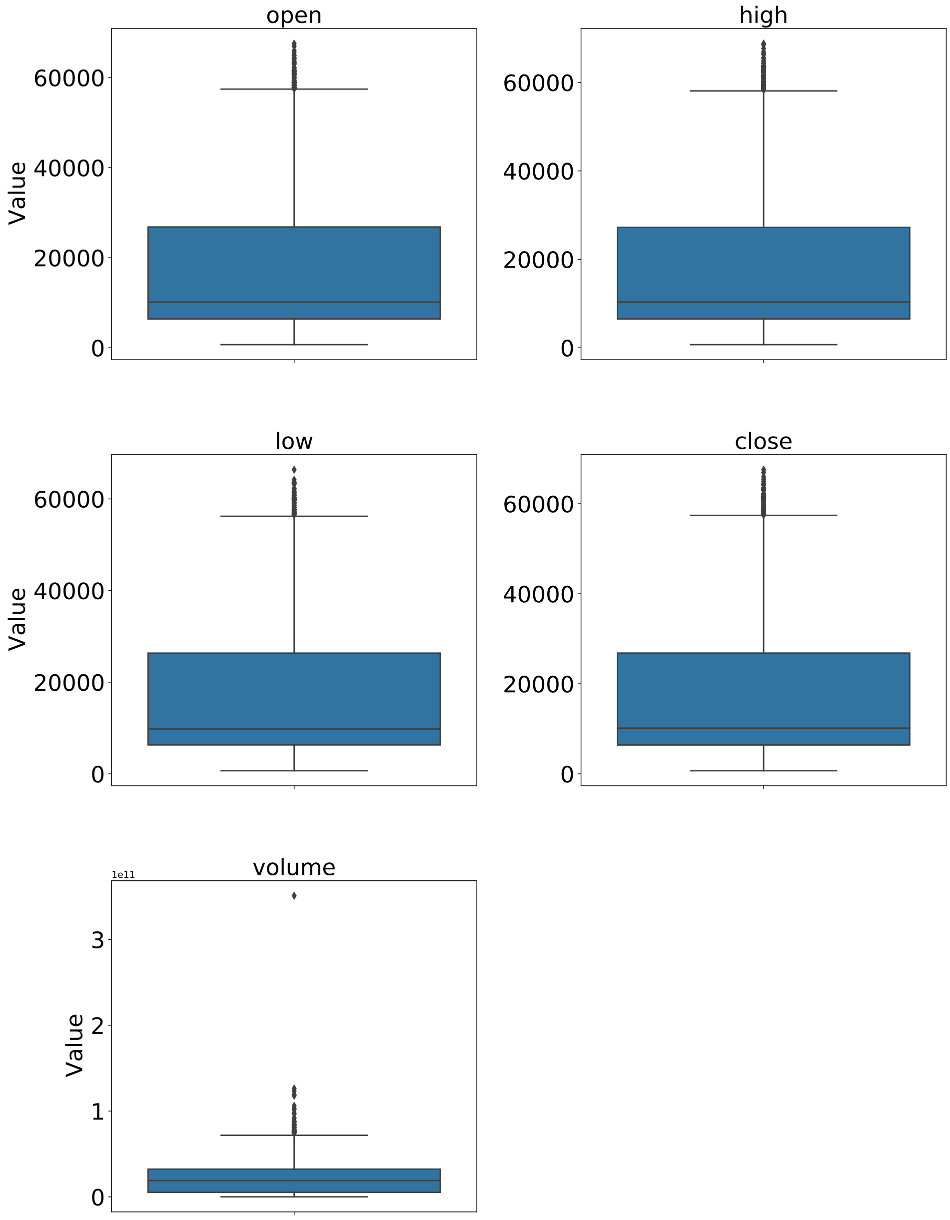

Figure 17: Box plots of the historical price features

APPENDIX B

Feature	**Granger Causality Observed From**	**Comments**
Volume Diff	Lag 2 onwards	No significant causality at lag 1; strong evidence from lags 2 to 12.
Daily Range Diff	Lag 3 onwards	No significant causality at lags 1 and 2; strong evidence from lags 3 to 12.
Price Change	Not significant	Most lags showed no significant causality, suggesting limited predictive power.
Twitter Sentiment Score	Lag 1 onwards	Strong causality observed from lag 1; p-values < 0.01.
Reddit Sentiment Score	Lag 1 onwards	Significant causality from lag 1; consistently low p-values.
Bitcointalk Sentiment Score	Lag 1 onwards	Very strong causality from lag 1; extremely low p-values (p < 0.0001).
Total Sentiment Score	Lag 1 onwards	Significant causality from lag 1; consistently low p-values.

Table 7: Results of the Granger Causality Test